1

Optimisation et réduction

<u>Optimisation</u> : recherche de la meilleure solution à un problème
<u>Réduction</u> : représentation simplifiée d'une donnée complexe

<u>Exemple typique</u> : réduction de dimension

Données en grande dimension, volonté de les représenter en 2D (e.g UMAP, TSNE)
Signaux représentés comme somme de signaux types Fourier

<u>LA QUESTION</u> que cherche-t-on à optimiser ?
Y-a-t-il une meilleure façon 'absolue' de trouver la meilleure réponse à une question ?
(Spoiler en général non)

I) <u>Un premier exemple : moindres carrés</u>

1) <u>Ajustement d'une courbe à partir de série temporelle</u>

<u>Cas 1 : la droite :</u>
On a des données (t_i, y_i), on cherche la meilleure droite :
$y = \alpha t + \beta$
qui minimise $\sum_{i=1}^{N} (\alpha t_i + \beta - y_i)^2$.

② On peut déterminer α et β en rappelant la chose suivante :
$\alpha, \beta \longrightarrow \sum (\alpha t_i + \beta - y_i)^2$ est une fonction régulière et au point de minimum on a

$$\frac{\partial}{\partial \alpha} = \frac{\partial}{\partial \beta} = 0$$

$$\frac{\partial}{\partial \alpha} = \sum (\alpha t_i + \beta - y_i) t_i$$

$$\frac{\partial}{\partial \beta} = \sum (\alpha t_i + \beta - y_i)$$

$$\Rightarrow \quad \begin{aligned} \alpha \left(\sum t_i^2\right) + \beta\left(\sum t_i\right) &= \sum y_i t_i \\ \alpha \sum t_i + \beta\left(\sum 1\right) &= \sum y_i \end{aligned}$$

C'est une équation !

$$M \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} u \\ v \end{pmatrix}$$

On peut généraliser ça aux paraboles

$$\alpha, \beta, \gamma \rightarrow \sum \left(\gamma t_i^2 + \alpha t_i + \beta - y_i\right)^2$$

<u>Exercice</u> : décrire la matrice

<u>NB</u> Compromis à faire entre nombre de données et paramètres.
N points et N paramètres = relier les points

③

Remarque 1: ce n'est pas tout à fait la distance du nuage de point à la droite.

Remarque 2: interprétation par maximum de vraisemblance

Modèle: $y = \alpha t + \beta + \underbrace{\mathcal{N}(0, \sigma^2)}_{\text{bruit}}$ σ^2 connue (erreur de mesure)

Densité de probabilité d'observer y_i au temps t_i pour des paramètres α, β

$$p(t_i, y_i \mid \alpha, \beta) = \frac{1}{\underset{\uparrow}{c(\sigma^2)}} e^{-\frac{(\alpha t_i + \beta - y_i)^2}{2\sigma^2}}$$

$\leftarrow$ renormalisation

Proba (densité d'observer toutes les données)

$$L(\alpha, \beta)\ p((t_1, y_1); (t_2, y_2) \dots) \mid \alpha, \beta) = \frac{1}{c(\sigma^2)^N} \prod_{i=1}^{N} e^{-\frac{(\alpha t_i + \beta - y_i)^2}{}}$$

On cherche à maximiser cela ie à minimiser $-\log L(\alpha, \beta)$

$$-\log L(\alpha, \beta) = + \underbrace{N \log c(\sigma^2)}_{\text{ind de } \alpha, \beta} + \frac{1}{\sigma^2} \sum_{i=1}^{N} (\alpha t_i + \beta - y_i)^2$$

$\Rightarrow$ $-\log L(\alpha, \beta)$ est minimale quand on minimise les moindres carrés.

II) Projection orthogonale

1) Principe

Rappel : distance euclidienne dans $\mathbb{R}^d$

$$d(x, y) = \sqrt{\sum_{i=1}^{d} (x_i - y_i)^2}$$

produit scalaire :

$$\langle \vec{x}, \vec{y} \rangle = \sum_{i=1}^{d} x_i \, y_i$$

on dit que

$$\vec{x} \perp \vec{y} \quad \text{si} \quad \langle \vec{x}, \vec{y} \rangle = 0$$

Espace vectoriel :

$0 \in F$

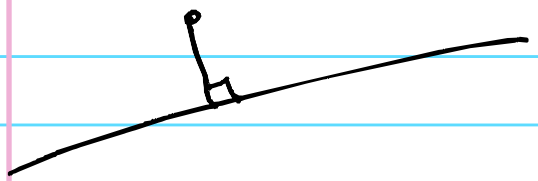
Si $u, v \in F$ $\qquad u + v \in F$

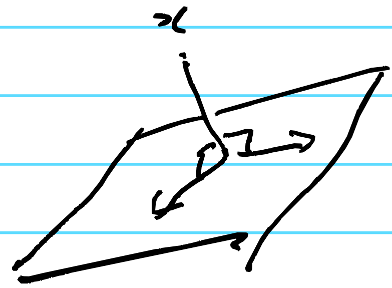
$\qquad \qquad \qquad \lambda u \in F$

(Espace affine) : Espace vectoriel shifté d'un point ("$x_0 + F$")

Projeté orthogonal de x sur un espace affine $(y_0 + F)$ est le point de $y_0 + F$: $x^\perp$ t.q : $x^\perp \in y_0 + F$

$$(x - x^\perp) \perp y \quad \forall \, y \in F$$





NB souvent pour simplifier on prend $x_0 = 0$.

5

2) Un premier cas : les séries de Fourier

Une première approche consiste à avoir des sous espaces définies à l'avance

On considère un signal T-périodique

<u>Idée</u> : Un signal periodique peut être vu comme une somme de sinusoïdes

$t \to \cos\left(2n\pi \frac{t}{T}\right)$ | $t \to e^{i2n\pi\frac{t}{T}}$

$t \to \sin\left(2n\pi \frac{t}{T}\right)$

version réelle

<u>Etape 1</u> : En quoi c'est orthogonal ?

<u>Produit scalaire</u>

réel $\langle f, g\rangle_{\mathbb{R}} = \frac{2}{T}\int_0^T f(s)\, g(s)\, ds$

complexe $\langle f, g\rangle_{\mathbb{C}} = \frac{1}{T}\int_0^T f(s)\, \overline{g(s)}\, ds$

<u>Exercice</u> : $\frac{2}{T}\int_0^T \cos\left(2n\pi \frac{s}{T}\right)\cos\left(2n'\pi \frac{s}{T}\right) ds = 0$ si $n \neq n'$, 1 si $n = n'$ ($n \neq 0$)

$\frac{2}{T}\int_0^T \sin\left(2n\pi \frac{s}{T}\right)\sin\left(2n'\pi \frac{s}{T}\right) ds = 0$, 1 si $n = n'$

$\frac{1}{T}\int_0^T \cos\left(2n\pi \frac{s}{T}\right)\sin\left(2n'\pi \frac{s}{T}\right) ds = 0$

$\frac{1}{T}\int_0^T e^{+2i n\pi \frac{s}{T}}\, e^{-i2i\pi n'\frac{s}{T}}\, ds = 0$ si $n \neq n'$, 1 sinon

⑥ $\boxed{\text{Question : quelle est la meilleure façon d'approcher une fonction (TT) périodique par des sinusoïdes de fréquence entre 0 et } N}$

ie si f est 2π-périodique, quelle est la meilleure fonction f_N ?

$$f_N(t) = \sum_{k=-N}^{N} c_k e^{ikt}$$

$$= \frac{a_0}{2} + \sum_{k=1}^{N} a_k \cos(kt) + b_k \sin(kt)$$

<u>Exercice</u> : donner les équations liant c_k, c_{-k} et (a_k, b_k)

Pourquoi la notion de meilleure est ambigüe : que cherche-t-on à rendre petit ?
Comment quantifier l'écart entre f et f_N ?

<u>Ecart max</u> $\quad \varepsilon_\infty(f, f_N) = \| f - f_N \|_\infty = \sup_{[0,T]} | f(t) - f_N(t) |$

<u>Ecart L^1</u> : $\quad \varepsilon_1(f, f_N) = \| f - f_N \|_1 = \int_0^T | f(t) - f_N(t) | \, dt$

<u>Ecart L^2</u> : $\quad \varepsilon_2(f, f_N) = \| f - f_N \|_2 = \sqrt{\int_0^T | f(t) - f_N(t) |^2 \, dt}$

(7) Avantage de ε_2 : calcul de ε_2^2

$$\frac{1}{2\pi}\int_0^{2\pi}\left|f(t)-\sum_{-N}^{N}\alpha_k e^{ikt}\right|^2 dt$$

$$=\frac{1}{2\pi}\int_0^{2\pi}\left(\overline{f(t)}-\sum_{-N}^{N}\bar\alpha_k e^{-ikt}\right)\left(f(t)-\sum_{-N}^{N}\alpha_k e^{ikt}\right)dt$$

$$=\frac{1}{2\pi}\int_0^{2\pi}\bar f(t) f(t)\,dt$$

$$-\sum_{-N}^{N}\alpha_k\frac{1}{2\pi}\int_0^{2\pi}\bar f(t)e^{ikt}dt-\sum_{-N}^{N}\bar\alpha_k\frac{1}{2\pi}\int_0^{2\pi}f(t)e^{-ikt}dt$$

$$+\sum_{k,k'}\bar\alpha_k\alpha_{k'}\frac{1}{2\pi}\int_0^{2\pi}e^{-ikt}e^{+ik't}dt$$

On note $c_k(f)=\frac{1}{2\pi}\int_0^{2\pi}f(t)e^{-ikt}dt$

$$\varepsilon_2^2\left(f,\sum_{-N}^{N}\alpha_k e^{ikt}\right)$$

$$=\frac{1}{2\pi}\int_0^{2\pi}|f(t)|^2dt$$

$$-\sum_{-N}^{N}\alpha_k\overline{c_k}(f)+\bar\alpha_k c_k(f)+c_k\overline{c_k}$$

$$=\frac{1}{2\pi}\int_0^{2\pi}|f(t)|^2-\sum_{-N}^{N}c_k(f)\,\overline{c_k}(f)$$

$$+\sum_{-N}^{N}(\alpha_k-c_k)(\bar\alpha_k-\bar c_k(f))$$

$$=\frac{1}{2\pi}\int_0^{2\pi}|f(t)|^2-\sum_{-N}^{N}|c_k(f)|^2 \to \text{indep de } \alpha$$

$$+\sum_{-N}^{N}|\alpha_k-c_k(f)|^2 \to \geq 0 \quad 0 \text{ ssi } \alpha_k=c_k$$

8

En quoi c'est une projection ?

$$f_N(t) = \sum_{-N}^{N} c_k(f)\, e^{ikt} \qquad g_N = \sum_{-N}^{N} c_k(g)\, e^{ikt}$$

$$\langle f - f_N , g_N \rangle$$

$$= \sum_{-N}^{N} \frac{1}{T}\int_0^T \left(\overline{f}(s) - \sum_{-N}^{N} \overline{c_k}\, e^{-iks} \right) \sum_{-N}^{N} c_{k'}(g) e^{ik's}\, ds$$

$$= \sum_{-N}^{N} \frac{1}{T}\int_0^T \overline{f}(s)\, e^{ik's} c_{k'}(g) - \overline{c_k}(f)\, c_k(g) e^{i(k'-k)s}\, ds$$

$$= \sum_{-N}^{N} \overline{c_{k'}}(f)\, c_{k'}(g) - \overline{c_k}(f)\, c_k(g)$$

$$= 0 \quad !$$

9

③ Adapter la base aux données : ACP

ACP : Analyse en Composantes Principales (PCA)

Question : On suppose qu'on a des données

X^j (j indice d'individus)

en dimension "grande"

$X^j = (X_i^j)_{1 \leq i \leq N >> 1}$.

Pour simplifier, on suppose qu'on a déjà centré les données ie

$$\sum_{(j)} X_i^j = 0 \quad \forall j$$

On voudrait représenter les données en 2D en gardant de la richesse.

Dispersion des données (variances)

$$\sum_j \sum_i (X_i^j)^2$$

Changement de base orthogonale

Pour le moment base $\underbrace{\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{\vec{e}_1}$ $\underbrace{\begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}}_{\vec{e}_2}$ $\dots$ $\underbrace{\begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}}_{\vec{e}_N}$

(10) Produit scalaire $\langle Y, Y' \rangle = \sum_1^N Y_i . Y'_i$

$Y \perp Y'$ $\qquad \sum_1^N Y_i Y'_i = 0$

$\langle \vec{e_i}, \vec{e_j} \rangle = 0$ si $i \neq j$

1 si $i = j$

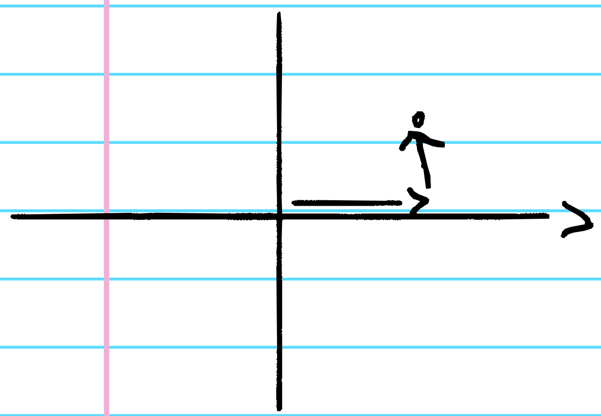
$\vec{v_1}, \ldots \vec{v_N}$ est une base orthonormée

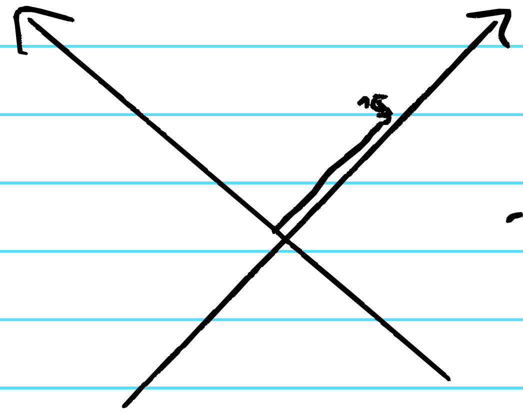
si $\langle \vec{v_i}, \vec{v_j} \rangle = 0$ si $i \neq j$

1 si $i = j$

Affirmation : $X^j = \sum X_i^j \vec{e_i}$

$= \sum \langle X^j, \vec{v_i} \rangle \vec{v_i}$





Propriété : $\sum_{i,j} (X_i^j)^2 = \sum \langle X^j, X^j \rangle$

$= \sum_j \sum_i (X_i^j)^2$

$= \sum_j \sum_i \langle X^j, \vec{v_i} \rangle^2$

11

Projection des X^j sur $\vec{e}_1, \vec{e}_2$

$\begin{pmatrix} X_1^j \\ X_2^j \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ on garde la proportion

$$\frac{\sum_j (X_1^j)^2 + (X_2^j)^2}{\sum_{i,j} (X_i^j)^2}$$

Projection des X^j sur $\vec{v}_1, \vec{v}_2$

$$\frac{\sum_j \langle X^j, \vec{v}_1 \rangle^2 + \langle X^j, \vec{v}_2 \rangle^2}{\sum_{i,j} (X_i^j)^2}$$

même

Peut on maximiser $\sum_j \langle X^j, \vec{v}_1 \rangle^2 + \langle X^j, \vec{v}_2 \rangle^2$?

Spoiler (oui)

Choix du premier vecteur

(complémenté)

(12)

II Mini introduction aux méthodes nonlinéaires

Comprendre le principe avec une fausse méthode

13

15